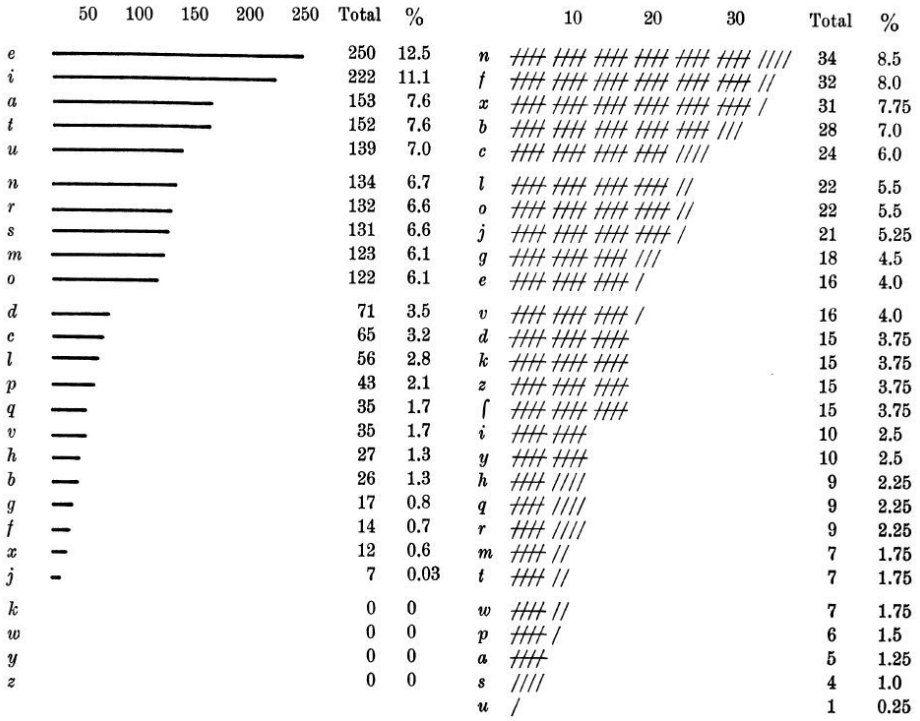
Le logogriphe d'Euler

Didier Müller, Lycée cantonal de Porrentruy

Le grand mathématicien suisse Leonhard **Euler** (1707-1783), extrêmement prolifique dans beaucoup de domaines, n'est pas connu pour avoir laissé son empreinte dans le domaine de la cryptographie. Cependant, en arithmétique, le théorème d'Euler sous-tend le cryptosystème RSA, largement utilisé dans les communications Internet. Mais ceci est une autre histoire…
À la fin d'une assez longue lettre à Christian Goldbach consacrée au calcul différentiel et datée du 4 juillet 1744, Euler propose un texte chiffré qu'il appelle « Logogryphum » et qui, à son avis, est difficile à décrypter.

pxqſwlznjdvynſtiddkqxhleebfpxdfgtlzbccfbkſodxokfnglqxnſchejmlckzxhrfwjgfh
xvzjnbgyxcdgixkoxjmlncoigdxvzflmeſnfyjqfangvnylrcxfonbfjalrkwſnbfpjoizoxq
knubroſadgiaxwkcbrbcklofrnjwgnſzfhgjfcbcfvqjtxeevtbzfyjsbzhſmlnbgſsqjwgl
nxvzfkonbcoigdxvrkfjalzxtſnilenfgvcboofcfxnnfgnkbcjnnjynxvplgnbfzfoxeejgdxbcj
cnſdyvdbhzlnvyxmbcblobbcyfekonbceiobfplwsxzxfjcndbhrlzqxsſonbcoljfſyqfm
jeevhleexoiexmgicfdnktvoldxnfbxoſcktvpxrnv

Il avait raison puisqu'il a fallu attendre 1953 pour que quelqu'un y parvienne ! Euler précise dans sa lettre que le texte clair est en latin, ce qui sera d'une grande aide. Reste à trouver le système de chiffrement...
C'est le suisse Pierre **Speziali** (1913-1995), chargé de cours en histoire des sciences à l'Université de Genève, qui a été le premier à publier la solution [1]. Hans **Rohrbach** (1903-1993), mathématicien et cryptanalyste allemand, l'a trouvée indépendamment peu après. Nous allons voir comment.
Rohrbach a d'abord fait une analyse des fréquences. L'image ci-dessous, tirée de son article [2], montre la comparaison entre les fréquences des lettres en latin (à gauche) et celles du logogriphe (à droite) :



Première constatation : le logogriphe contient les 26 lettres de notre alphabet plus le signe *ſ*, une sorte de *S* allongé qui ressemble à un « *f* » sans tiret. L'alphabet latin, lui, ne compte que 20 lettres dans sa variante archaïque. Rohrbach a différencié dans ses statistiques le I du J et le U du V, c'est pourquoi son alphabet latin compte deux lettres de plus.
Les cinq lettres latines les plus fréquentes (E, I, A, T, U) représentent 45 % d'un texte. Dans le logogriphe, qui compte 408 lettres, elles devraient donc apparaître environ 37 fois chacune. Or, le *n*, qui revient le plus souvent, n'y figure que 34 fois. Cela pourrait signifier qu'Euler a employé au moins deux représentations pour les lettres les plus fréquentes, ce qui met sur la piste d'un chiffre *homophonique* : une lettre peut être chiffrée par plusieurs symboles différents, appelés *homophones*. Par chance, comme il n'y a que six lettres et le signe *ſ* en supplément, cela suggère qu'il y a peu d'homophones : probablement sept (27-20), mais cela reste à confirmer.
Rohrbach examine alors des groupes de lettres qui apparaissent le plus fréquemment dans le texte [2] :

	a	b	c	d	e	f	g	h	i	j
1	*pxqſw*	*lznjd*	*vynſt*	*iddkq*	*xhlee*	*bfpxd*	*fgtlz*	*bccfb*	*kſodx*	*okfngl*
2	*qxnſc*	*hejml*	*ckzxh*	*rfwjg*	*fhxvz*	*jnbgy*	*xcdgi*	*xkoxj*	*mlnco*	*igdxvz*
3	*flmeſ*	*nfyjq*	*fangv*	*nylrc*	*xfonb*	*fjalr*	*kwſnb*	*fpjoi*	*zoxqk*	*nubroſ*
4	*adgia*	*xwkcb*	*rbckl*	*ofrnj*	*wgnſz*	*fhgjf*	*cbcfv*	*qjtxe*	*evtbz*	*fyjsbz*
5	*hſmln*	*bgſsq*	*jwgln*	*xvzfk*	*onbco*	*igdxv*	*rkfja*	*lzxtſ*	*nilen*	*fgvcbo*
6	*ofcfx*	*nnfgn*	*kbcjn*	*njynx*	*vplgn*	*bfzfo*	*xeejg*	*dxbcj*	*cnſdy*	*vdbhzl*
7	*nvyxm*	*bcblo*	*bbcyf*	*ekonb*	*ceiob*	*fplws*	*xzxfj*	*cndbh*	*rlzqx*	*sſonbc*
8	*oljff*	*yqfmj*	*eevhl*	*eexoi*	*exmgi*	*cfdnk*	*tvold*	*xnfbx*	*oſckt*	*vpxrnv*

Il remarque que la suite *coigdxv* (aux lignes 2 et 5) est suivie une fois par *z* et l'autre par *r*. C'est aussi le cas pour la suite *fjal* (aux lignes 5 et 3). Il suppose donc que *r* et *z* sont des homophones. Cela semble être confirmé par le groupe de 5 lettres *dbhzl* (ligne 6, colonne j) comparé à *dbhrl* (ligne 7, colonnes h et i).
En procédant ainsi, il trouve les cinq paires d'homophones *j-f*, *k-x*, *w-t*, *y-g* et *z-r*.
Il peut maintenant réécrire le logogriphe en éliminant les homophones, et il souligne à nouveau les répétitions, qui sont plus nombreuses [2] :

pxqſt	*lrnfd*	*vgnſt*	*iddxq*	*xhlee*	*bfpxd*	*fgtlr*	*bccfb*	*xſodx*	*oxfngl*
qxnſc	*hefml*	*cxrxh*	*rftfg*	*fhxvr*	*fnbgg*	*xcdgi*	*xxoxf*	*mlnco*	*igdxvr*
flmeſ	*nfgfq*	*fangv*	*nglrc*	*xfonb*	*ffalr*	*xtſnb*	*fpfoi*	*roxqx*	*nubroſ*
adgia	*xtxcb*	*rbcxl*	*ofrnf*	*tgnſr*	*fhgff*	*cbcfv*	*qftxe*	*evtbr*	*fgfsbr*
hſmln	*bgſsq*	*ftgln*	*xvrfx*	*onbco*	*igdxv*	*rxffa*	*lrxtſ*	*nilen*	*fgvcbo*
ofcfx	*nnfgn*	*xbcfn*	*nfgnx*	*vplgn*	*bfrfo*	*xeefg*	*dxbcf*	*cnſdg*	*vdbhrl*
nvgxm	*bcblo*	*bbcgf*	*exonb*	*ceiob*	*fplts*	*xrxff*	*cndbh*	*rlrqx*	*sſonbc*
olfff	*gqfmf*	*eevhl*	*eexoi*	*exmgi*	*cfdnx*	*tvold*	*xnfbx*	*oſcxt*	*vpxrnv*

Ensuite, Rohrbach examine les fréquences des *bigrammes* (groupes de deux lettres successives). Pour ce faire, il réalise le tableau statistique ci-après. Il ne compte pas uniquement les fréquences des lettres, mais note aussi la lettre qui la suit dans le logogriphe.
Par exemple, sur la première ligne du tableau, il note que le *f* est suivi du *g* 7 fois, du *f* 4 fois, du *c* 3 fois, etc. Le *f* apparaît en tout 53 fois, ce qui représente 12.99 % des lettres.

		abs.	%
f	*g*:::. *f*:: *c*:. *t*:. *a*: *m*: *n*: *p*: *o*: *d*. *b*. *h*. *r*. *e*. *x*. *l q v s ſ*	53	12.99
x	*o*::. *f*:: *v*:: *t*:: *n*:. *q*: *r*: *h*. *c*. *b*. *e*. *m*. *d x l s ſ*	46	11.28
n	*b*:::. *f*::: *x*:. *ſ*:. *g*: *n*. *v*. *c d i u*	34	8.33
b	*e*:::::. *f*:: *r*:. *g*. *h*. *x*. *b l o*	28	6.86
g	*f*:: *n*:: *v*: *d*: *l*: *i*: *x*. *g q t ſ*	28	6.86
c	*f*::. *b*:. *x*:. *o*: *n*. *c h d e g*	24	5.88
r	*f*::: *x*:: *n*: *b*. *o*. *l*. *c h q*	24	5.88
l	*r*::. *n*:. *e*: *o*. *c d g m f q*	22	5.39
o	*x*:. *n*:. *i*:. *b*. *f*. *l*. *ſ*. *d*	22	5.39
e	*e*:: *x*: *f*. *v*. *b i n ſ*	16	3.92
v	*r*: *g*. *p*. *c d o n h t q*	16	3.92
d	*x*::. *g*: *b*. *d f n v*	15	3.67
ſ	*n*: *t*. *o*. *c*. *r d g m a s*	15	3.67
t	*x*. *g*. *l*. *ſ*. *v*. *f i b s*	14	3.43
i	*g*. *r c d o a x l e*	10	2.45
h	*r*: *l*. *x e g ſ*	9	2.20
q	*f*:. *x*:. *ſ*	9	2.20
m	*l*: *b e f g*	7	1.71
p	*x*: *l*. *f*	6	1.47
a	*l*. *d n x*	5	1.22
s	*x b q ſ*	4	0.98
u	*b*	1	0.24
j	—		
k	—		
w	—		
y	—		
z	—		

En latin, la lettre Q est assez courante et vient toujours juste avant un U. Rohrbach cherche donc un bigramme dont la première lettre n'apparaît qu'en combinaison avec une certaine autre. Or, cela n'arrive qu'une seule fois : *u* suivi de *b*. C'est trop peu. Il conclut qu'Euler a probablement utilisé une seule lettre pour chiffrer « QU ».
Il n'y a que quatre lettres qui peuvent venir après QU : A, E, I et O. Toutes les quatre sont assez communes et plus fréquentes que Q. Les candidats sont donc : *q*, *p*, *a* et *s*.
Fort de ces constatations, Rohrbach finit par identifier que *p* = QU, *f* = E, *x* = I et *l* = A.
Un autre point marquant dans le tableau ci-dessus est le bigramme « *ee* » qui apparaît 4 fois. Il doit correspondre à un redoublement de consonnes dans le texte clair. Une lettre peu fréquente qui se redouble souvent en latin est le L. Pour en être sûr, il faut retrouver les occurrences de « *ee* » dans le logogriphe [2] :

	i all equi			*e il l*			*ille*		*e e ll a lli*
1e,f	*xhlee bfp*	*xd*	4h,i	*qftxe evtbr*	6g	*xeefg*	8b,c,d	*gqfmf eevhl eexoi*	

Ainsi donc, *ee* correspond bien à LL.
Dans la partie 8b,c,d, il est relativement aisé de deviner les lettres manquantes et l'on voit apparaître « de bello gallico » [2].

	de b e ll oga ll ico	
8b,c,d	*gqfmf eevhl eexoi*	

Le reste est de la routine pour un cryptanalyste. À l'aide des lettres déjà découvertes, des mots et donc d'autres lettres peuvent être devinées [2] :

caesa	*rdeb e lloga*	*ll ico*	*li b r o*	*septi*	*mocap*	*i t ev i*	*cesim*	*oquinto*	
olffl	*gqfmf*	*eeihl*	*eexoi*	*exmgi*	*cfdnx*	*t iold*	*xnfbx*	*olcxt*	*ip xrni*
?									

De proche en proche, Rohrbach détermine la table de chiffrement suivante :

Clair	A	B	C	D	E	F	G	H	I, J	L	M	N	O	P	QU	R	S	T	U, V	X
Chiffré	*l*	*m*	*o*	*q*	*j*	*s*	*h*	*u*	*k*	*e*	*t*	*r*	*i*	*d*	*p*	*g*	*c*	*n*	*b*	*a*
	f				*f*				*x*		*w*	*z*	*v*			*y*	*f* ?			

On peut constater qu'Euler a utilisé des homophones pour les lettres claires A, E, I/J, M, N, O et R. **Speziali** fait remarquer dans son article que « le *f* du chiffre signifie tantôt E, tantôt S, ce qui veut dire que dans le manuscrit original, aujourd'hui introuvable, il y avait deux *f* de type différent. C'est probablement l'une des causes des difficultés qu'a présentées le décryptement de ce texte » [1]. La deuxième possibilité est qu'Euler s'est trompé lors de son chiffrement…
Le texte déchiffré est un extrait de « La Guerre des Gaules » de Jules César (livre VII, chapitre 25) :

> *Quidam ante portam oppidi Gallus, qui per manus sebi ac picis traditas glebas in ignem e regione turris proiciebat, scorpione ab latere dextro traiectus exanimatusque concidit. Hunc ex proximis unus iacentem transgressus eodem illo munere fungebatur ; eadem ratione ictu scorpionis exanimato alteri successit tertius et tertio quartus, nec prius ille est a propugnatoribus vacuus relictus locus quam restincto aggere atque omni ex parte summotis hostibus finis est pugnandi factus.*
> *Caesar, De bello gallico, libro septimo, capite vicesimo quinto*
>
> Devant la porte de la ville était un Gaulois, à qui l'on passait de main en main des boules de suif et de poix, qu'il lançait dans le feu du haut d'une tour. Un trait de scorpion lui perce le flanc droit ; il tombe mort. Un de ses plus proches voisins passe par-dessus le cadavre et remplit la même tâche ; il est atteint à son tour et tué de la même manière ; un troisième lui succède ; à celui-ci un quatrième ; et le poste n'est abandonné que lorsque le feu de la terrasse est éteint et que la retraite des ennemis partout repoussés a mis fin au combat.
> César, La Guerre des Gaules, livre VII, chapitre 25

Notons pour finir que la méthode de décryptement de Speziali est différente, bien que basée elle aussi sur les fréquences d'apparition des lettres.
Lui s'est d'abord attaqué aux lettres redoublées afin de faire la distinction entre consonnes et voyelles. Les voici dans l'ordre, avec les deux lettres qui les encadrent : *iddk, leeb, bccf, xeev, boof, xnnf, jnnj, xeej, obbc, jeev* et *leex.* Puisque toutes les lettres qui encadrent les redoublements ne se redoublent pas elles-mêmes, il en tire que *i*, *k, l, f, x, v* et *j* sont des voyelles. Le *b* fait exception, mais comme il encadre trois redoublements et comme *o* et *c* se redoublent en tant que consonnes, il n'y a aucun risque à supposer que *b* est une voyelle.
Puis il porte son effort sur les lettres rares du cryptogramme, car elles correspondent probablement à des lettres rares du texte clair. Le *a* n'apparaît que 5 fois et précisément dans les groupes *fangv, jalrk, fadgiaxwkc, jalzx.* Il utilise les trigrammes *ang* et *adg* pour déterminer *a.* Après deux essais seulement, son choix se porte sur la lettre X. Dès lors, les possibilités se réduisent à une seule, *ang* = XTR, ce trigramme étant bien encadré par deux voyelles. La voyelle qui le précède est certainement *e,* donc *fang* = EXTR. Il pose ensuite *adg* = XPR, puis trouve que probablement *fadgiaxwkc* = EXPROXIMIS. Avec les lettres ainsi découvertes, il réussit à décrypter petit à petit tout le logogriphe.

Références

[1] Speziali Pierre, « Le logogriphe d'Euler », Stultifera navis, Bull. Soc. suisse des bibliophiles, avril 1953, pp. 6-9
[2] Rohrbach Hans, « The logogryph of Euler », Journal für die reine und angewandte Mathematik. Band 262/263, 1973, pp. 392-399

Ces deux articles, ainsi que la lettre d'Euler à Goldbach de 1744, sont disponibles sur cette page web :
www.apprendre-en-ligne.net/crypto/euler/